

Penerapan Algoritma Naive Bayes Classifier untuk Analisis Sentimen Komentar Twitter Proyek Pembangunan IKN

Faiz Zamzami¹, Rahmat Hidayat², Rina Fathonah³

^{1,3}Department of Data Science, Universitas Putra Bangsa, Indonesia

²Department of Informatic, Universitas Putra Bangs, Indonesia

Article Info

Article history:

Received Feb 2, 2024

Revised Mar 18, 2024

Accepted Mar 26, 2024

Keywords:

Data Mining

IKN

Naïve Bayes Classifier

Sentiment Analysis

Twitter

ABSTRACT

Recently, the relocation of Indonesia's capital city has become a hot topic of discussion among the public. Various opinions emerged regarding this mega project proposed by President Joko Widodo. Especially on social media, Twitter has become one of the most popular platforms in Indonesia as a forum for expressing people's opinions in public. In the context of the development of IKN Nusantara, researchers conducted an analysis of Twitter users' comments on President Joko Widodo's official account. Using the Naïve Bayes method with a dataset containing 220 comments consisting of 116 negative comments, 70 positive comments and 34 neutral comments. In this research, researchers developed a Python-based machine learning program. The analysis results show respective values of Precision 62%, Recall 66%, and f1-Score 63% with an accuracy level of 66%. In testing using 20% of the data, the program successfully predicted 20 negative comments, 8 neutral comments, and 16 positive comments.

Copyright © 2024 Universitas Indraprasta PGRI.
All rights reserved.

Corresponding Author:

Faiz Zamzami,

Department of Data Science,

Universitas Putra Bangsa,

Jl. Ronggowarsito No.18, Kedawung, Pejagoan, Kebumen, Jawa Tengah.

Email: faizzamzami10p@gmail.com

1. PENDAHULUAN

Semakin meningkatnya pembicaraan di kalangan masyarakat tentang pemindahan Ibu Kota Indonesia, banyak orang mengemukakan pendapat mereka sebagai cara untuk mengekspresikan diri. Lalu akan di pindah dimana Ibukota baru Indonesia? Ibu kota Indonesia (IKN) akan dipindahkan ke Kabupaten Penajam Paser Utara dan Kabupaten Kutai Kartanegara di Provinsi Kalimantan Timur. Beberapa kriteria yang menjadi pertimbangan pemilihan Kalimantan Timur untuk IKN baru meliputi lokasi yang mudah diakses, kedekatan dengan dua kota besar yaitu Balikpapan dan Samarinda, struktur kependudukan yang heterogen dan terbuka, potensi konflik yang rendah, infrastruktur utama yang lengkap dengan dukungan bandar udara dan Pelabuhan [1]. Lalu kenapa harus ada pemindahan Ibukota? Jakarta menghadapi beban yang semakin berat sebagai pusat pemerintahan dan ekonomi. Jakarta yang awalnya direncanakan untuk menampung 600 ribu jiwa pada era kolonial, kini dihuni oleh 10 juta penduduk yang mengakibatkan berbagai permasalahan sosial seperti polusi, kemacetan, banjir, dan kemiskinan yang semakin meresap dalam struktur kota ini [1]. Isu tersebut menjadi semakin jelas dengan disahkannya RUU IKN oleh pemerintah dan DPR pada bulan Januari 2022 yang semakin menegaskan komitmen pemerintah untuk memindahkan Ibukota negara ke pulau Kalimantan.

Lalu bagaimana tanggapan masyarakat mengenai pemindahan Ibukota ke Kalimantan? Berlandaskan pertanyaan tersebut di lakukan penelitian ini untuk mengetahui pendapat dan opini-opini masyarakat mengenai IKN Nusantara. Penelitian ini memanfaatkan perkembangan teknologi semakin hari semakin maju dan berkembang pesat terutama internet. Pertumbuhan pesat internet dan media sosial menyediakan beragam informasi dan pendapat dari berbagai individu [2]. Perkembangan teknologi informasi dapat dilihat dengan semakin banyaknya media sosial bermunculan seperti YouTube, Facebook, Instagram, Twitter, Tik Tok dll. Salah satu media sosial yang paling banyak digunakan adalah Twitter. Twitter adalah platform media sosial yang memfasilitasi pengguna untuk mengirim dan membaca pesan berbentuk teks dengan batasan maksimal 140 karakter yang biasa disebut Tweet [3]. Tweet dapat berupa sebuah persepsi dan opini dari para pengguna untuk menanggapi hal-hal yang sedang *Tren* saat itu. Dengan jumlah pengguna sebanyak itu memungkinkan banyaknya data tweet yang ada setiap hari. Dari data tweet tersebut dapat dijadikan sebagai data untuk Analisis Sentimen. Ada beberapa potensi reaksi yang mungkin timbul dari masyarakat, termasuk respons yang bersifat positif, negatif, dan netral yang dapat diidentifikasi secara langsung melalui tulisan-tulisan yang dipublikasikan oleh masyarakat di media sosial [4].

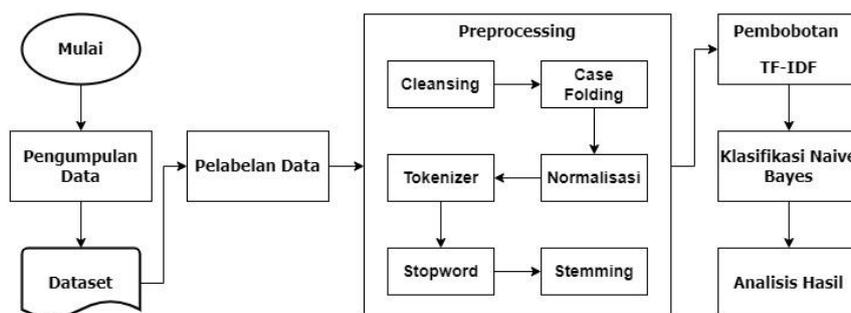
Untuk menganalisis sentimen dari pendapat masyarakat, pengklasifikasian dapat dilakukan menggunakan algoritma Naive Bayes Classifier. Naive Bayes Classifier adalah algoritma klasifikasi yang menggunakan frekuensi kata dalam dokumen untuk menghitung probabilitas sebuah kelas atau label menggunakan Teorema Bayes [5]. Algoritma Naive Bayes berfungsi sebagai sebuah metode untuk mengidentifikasi pola-pola kesamaan karakteristik dalam sebuah kelas atau label data tertentu [6]. Pemilihan algoritma Naive Bayes disebabkan kemampuannya dalam memproses data dengan kecepatan tinggi dan memberikan tingkat akurasi yang signifikan, terutama saat diterapkan pada dataset yang bersifat besar, beragam, dan melibatkan jumlah data yang besar [7]. Penentuan polaritas positif atau negatifnya suatu opini dapat dilakukan secara manual, tetapi seiring bertambahnya sumber opini menjadi semakin banyak tentunya waktu dan usaha yang dibutuhkan untuk mengklasifikasikan polaritas opini tersebut akan semakin banyak terpakai. Oleh karena itu, diajukan penerapan metode pembelajaran mesin untuk mengklasifikasi polaritas opini dari sumber data yang sangat banyak tersebut. Untuk melakukan hal itu, bisa menggunakan salah satu fungsi dari text mining dalam hal ini adalah Analisis Sentimen [8]. Pengertian Text Mining di sini mencakup langkah-langkah proses otomatisasi dalam analisis teks yang dilakukan oleh komputer untuk mengeksplorasi informasi berkualitas dari serangkaian teks yang tercakup dalam sebuah dokumen [9].

Sudah banyak penelitian sebelumnya yang menggunakan algoritma Naive Bayes sebagai metode pembelajaran mesin untuk melakukan Analisis Sentimen. Salah satunya penelitian yang dilakukan oleh Rizqia Lestika Atimi dengan topik tentang “Implementasi Model Klasifikasi Sentimen Pada Review Produk Lazada Indonesia”. Dari penelitian tersebut didapatkan hasil penerapan algoritma Naive Bayes pada 313 data uji, terbukti bahwa model yang dibuat mampu mengategorikan semua ulasan produk Lazada Indonesia ke dalam tiga kelas, yaitu 116 positif, 101 negatif, dan 96 netral. Dari hasil tersebut dapat disimpulkan bahwa 32,27% dari konsumen Lazada Indonesia merasa puas dengan performa penjual dan produk yang dijual di platform tersebut [10].

Pada penelitian lainnya yang dilakukan oleh Mutiara Azahri dengan judul “Analisis Sentimen Pengguna Kereta Api Indonesia Melalui Sosial Media Twitter Dengan Algoritma Naive Bayes Classifier”. Dari hasil penelitian menggunakan algoritma Naive Bayes Classifier berhasil mencapai akurasi sebesar 92.15% dalam mengklasifikasikan sentimen. Hasil menunjukkan bahwa layanan transportasi Kereta Api Indonesia mendapat sentimen negatif, terutama dari pengguna Twitter yang mengeluh tentang harga tiket, kursi, dan fasilitas lainnya [11].

2. METODE

Dalam penelitian ini, peneliti berusaha menganalisis komentar pengguna Twitter yang menyatakan pendapat mereka mengenai tweet akun Bapak Joko Widodo terkait mega proyek pembangunan Ibu Kota Negara (IKN) di Kalimantan Timur. Tujuan utamanya adalah untuk memahami bagaimana tanggapan masyarakat, terutama pengguna Twitter, terhadap mega proyek IKN. Dari total 200 data yang diambil sebagai sampel, 20% akan dialokasikan sebagai data uji dan 80% sebagai data latihan. Proses analisis sentimen akan dilakukan menggunakan aplikasi web Jupyter Notebook sebagai editor kode dengan menggunakan bahasa pemrograman Python. Dalam dunia pemrograman menggunakan bahasa Python, terdapat berbagai library dan framework yang dapat dimanfaatkan untuk melakukan analisis data [12][3]. Library yang digunakan antara lain Pandas, NLTK, Sastrawi, Wordcloud, Sklearn dan Imlearn. Untuk langkah-langkah metode penelitian Analisis Sentimen pada penelitian ini dijabarkan sebagai berikut:



Gambar 1. Alur Analisis Sentimen Naive Bayes

2.1 Pengumpulan Data

Data penelitian ini diperoleh melalui analisis komentar yang terdapat pada akun Twitter Bapak Joko Widodo, terkait dengan isu IKN (Ibu Kota Negara). Komentar tersebut mencakup aspek-aspek seperti pemerataan ekonomi, pemerataan penduduk, dan penciptaan titik-titik pertumbuhan ekonomi baru terkait dengan konsep Pembangunan Ibu Kota Negara yang terletak di pulau Kalimantan. Proses pengumpulan data melibatkan pemilihan komentar yang memenuhi beberapa kriteria, antara lain tidak mengandung unsur politik atau kampanye, menggambarkan pendapat penulis dengan jelas, dan tidak mengandung unsur rasisme. Sebanyak 220 komentar yang memenuhi kriteria tersebut diambil sebagai sampel data untuk dilakukan analisis sentimen.

2.2 Pelabelan Data

Penentuan label pada data hasil identifikasi dilakukan secara manual dengan menyesuaikan opini terhadap sentimen positif dan negatif. Sentimen positif diberikan pada data yang mengandung kalimat atau opini yang positif serta mendukung, sedangkan sentimen negatif diberikan pada data yang mengandung kalimat atau opini yang negatif serta merendahkan [13]. Sedangkan, tweet akan mendapatkan label netral jika jumlah kata positif dan negatif sama atau bernilai 0.

2.3 Tahap Preprocessing

Tahap pre-processing adalah langkah yang dijalankan sebelum penerapan metode klasifikasi. Fungsinya adalah menghapus noise, menormalisasi bentuk kata, dan mengurangi jumlah kosakata [14], [15]. Tahapan preprocessing melibatkan beberapa langkah yaitu Pembersihan (Cleansing), Penyederhanaan Huruf (Case Folding), Pembagian Kata (Tokenizer), Normalisasi Bahasa, Penyaringan (Stopword), Pemangkasan Akar (Stemming).

2.4 Proses Pembobotan Kata TF-IDF

Merupakan suatu proses untuk mengubah data dari bentuk teks menjadi bentuk numerik, yang kemudian dapat dijadikan dasar untuk memberikan bobot untuk menilai seberapa signifikan suatu kata dalam suatu dokumen [16]. Dengan memberikan bobot pada setiap kata yang mencerminkan seberapa sering dan seberapa signifikan kata tersebut muncul dalam suatu data atau dokumen.

2.5 Klasifikasi Naive Bayes Classifier

Naive Bayes Classifier adalah algoritma klasifikasi yang menggunakan frekuensi kata dalam dokumen untuk menghitung probabilitas sebuah dokumen menggunakan Teorema Bayes. Metode Naive Bayes sangat terkenal karena kepraktisannya. Keunggulan dari metode ini terletak pada kemampuannya untuk memenuhi kebutuhan dalam berbagai domain yang berbeda [17]. Dalam penelitian ini sistem juga dirancang untuk dapat memprediksi komentar yang baru di tambahkan dan mengklasifikasikan dengan label positif, netral, ataupun negatif tergantung kata yang terdapat pada komentar baru. Dasar dari implementasi Naive Bayes dalam pemrograman adalah menggunakan rumus Bayes. Rumus ini menggambarkan peluang kejadian A sebagai B, yang ditentukan oleh peluang B ketika A terjadi, peluang A, dan peluang B [15]. Contohnya pada persamaan beriku ini:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

Keterangan:

- a) $P(H|X)$ = Peluang hipotesa H berdasarkan kondisi X

- b) $P(X|H)$ = Peluang X berdasarkan kondisi hipotesa H
- c) $P(H)$ = Peluang hipotesa H
- d) $P(X)$ = Peluang dari X yang diamati
- e) H = hipotesa X data dengan label
- f) X = data sampel dengan label yang tidak diketahui

2.6 Analisis Hasil

Pada bagian analisis hasil akan ada beberapa hasil analisis yang akan dipaparkan. Diantaranya analisis akurasi dari penggunaan Naïve Bayes sebagai algoritma machine learning yang digunakan pada sentimen analis, hasil prediksi precision, recall, dan f1-score, penggambaran prediksi model Naïve Bayes menggunakan confusion matrix, dan analisis dan visualisasi per-kata menggunakan Wordcloud.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Pengumpulan data pada penelitian ini menggunakan platform media sosial Twitter sebagai sumber data untuk mencari komentar masyarakat terakait pembangunan IKN. Data di ambil dari tweet akun Bapak Joko Widodo pada bulan September sampai Desember 2023. Data yang di kumpulkan berupa username dan teks komentarnya. Tabel 1 berikut ini merupakan contoh sampel data.

Tabel 1. Ilustrasi Sampel Dataset

Teks Komentar
"@dAsalbantani~IKN menjadi langkah strategis untuk menciptakan pemerataan ekonomi dan penduduk, menjauhkan fokus dari Pulau Jawa. Pembangunan ini membutuhkan kesabaran, tetapi berpotensi memberikan dampak positif jangka panjang.
"@LengkengMasak~ Kenapa duit segitu gede tidak di bagi ke beberapa kota untuk pemerataan Ekonomi, IKN udahlah biaya besar dan terfokus ke satu titik.
"@AnugrahPrahasta~Prioritaskan pekerjaan untuk rakyat pak. Tanpa rakyat yang bekerja, negaranya bisa lumpuh.
"@PasukanPelor~Kita gak butuh IKN pak, yg kita butuh sekolah2 dan fasilitas kesehatan yg merata di daerah2.
"@abdulmukti691~Mending dana triliunan utk IKN digunakan utk subsidi pupuk, subsidi listrik, dan buat bangun sekolah. IKN itu projek gak berguna

3.2 Pelabelan Data

Proses pelabelan data melibatkan identifikasi data mentah dan penambahan satu atau beberapa label bermakna. Dalam penelitian ini, pemberian label pada tweet dilakukan secara manual, dengan kategori sentimen positif, negatif, dan netral. Menurut penulis, sentimen positif diberikan pada data yang mengandung kalimat atau opini positif dan membangun, sementara sentimen negatif diberikan pada data yang berisi kalimat atau opini negatif dan merendahkan. Sebaliknya, tweet akan mendapatkan label netral jika jumlah kata positif dan negatif sama atau bernilai 0. Hasil dari proses pelabelan data dapat dilihat pada Tabel berikut.

Tabel 2. Ilustrasi Pelabelan Data

Komentar	Sentimen
"@dAsalbantani~IKN menjadi langkah strategis untuk menciptakan pemerataan ekonomi dan penduduk, menjauhkan fokus dari Pulau Jawa. Pembangunan ini membutuhkan kesabaran, tetapi berpotensi memberikan dampak positif jangka panjang.	Positif
"@LengkengMasak~ Kenapa duit segitu gede tidak di bagi ke beberapa kota untuk pemerataan Ekonomi, IKN udahlah biaya besar dan terfokus ke satu titik.	Negatif
"@AnugrahPrahasta~Prioritaskan pekerjaan untuk rakyat pak. Tanpa rakyat yang bekerja, negaranya bisa lumpuh.	Netral
"@PasukanPelor~Kita gak butuh IKN pak, yg kita butuh sekolah2 dan fasilitas kesehatan yg merata di daerah2.	Negatif

"@abdulmukti691~Mending dana triliunan utk IKN digunakan utk subsidi pupuk, subsidi listrik, dan buat bangun sekolah. IKN itu projek gak berguna

3.3 Tahap Preprocessing

Pada tahap ini data akan diubah kedalam format yang sesuai dengan kebutuhan untuk mempermudah proses pengklasifikasian, sehingga menghasilkan hasil yang maksimal.

- a. Pembersihan (Cleansing): Tahap pertama yang dilakukan untuk membersihkan data dari simbol dan link URL yang tidak dibutuhkan. Tabel 3 berikut ini merupakan hasil dari proses pembersihan.

Tabel 3. Ilustrasi Cleansing Data

Cleansing
<p>IKN menjadi langkah strategis untuk menciptakan pemerataan ekonomi dan penduduk menjauhkan fokus dari Pulau Jawa Pembangunan ini membutuhkan kesabaran tetapi berpotensi memberikan dampak positif jangka panjang Kenapa duit segitu gede tidak di bagi ke beberapa kota untuk pemerataan Ekonomi IKN sudahlah biaya besar dan terfokus ke satu titik Prioritaskan pekerjaan untuk rakyat pak Tanpa rakyat yang bekerja negaranya bisa lumpuh Kita gak butuh IKN pak yg kita butuh sekolah2 dan fasilitas kesehatan yg merata di daerah2 Mending dana triliunan utk IKN digunakan utk subsidi pupuk subsidi listrik dan buat bangun sekolah IKN itu projek gak berguna</p>

- b. Case Folding: Setelah teks data dibersihkan, tweet yang telah dibersihkan akan diubah menjadi huruf kecil atau huruf cetak kecil. Hasilnya dapat di lihat pada Tabel 4.

Tabel 4. Ilustrasi Case Folding

Case Folding
<p>ikn menjadi langkah strategis untuk menciptakan pemerataan ekonomi dan penduduk menjauhkan fokus dari pulau jawa pembangunan ini membutuhkan kesabaran tetapi berpotensi memberikan dampak positif jangka panjang kenapa duit segitu gede tidak di bagi ke beberapa kota untuk pemerataan ekonomi ikn sudahlah biaya besar dan terfokus ke satu titik prioritaskan pekerjaan untuk rakyat pak tanpa rakyat yang bekerja negaranya bisa lumpuh kita gak butuh ikn pak yg kita butuh sekolah2 dan fasilitas kesehatan yg merata di daerah2 mending dana triliunan utk ikn digunakan utk subsidi pupuk subsidi listrik dan buat bangun sekolah ikn itu projek gak berguna</p>

- c. Normalisasi: Tahap mengubah bahasa informal (slang words) atau kata-kata tidak baku menjadi kata-kata baku. Tabel 5 berikut ini merupakan hasil dari proses tokenizer atau pemisahan kata.

Tabel 5. Ilustrasi Normalisasi

Normalisasi
<p>ikn menjadi langkah strategis untuk menciptakan pemerataan ekonomi dan penduduk menjauhkan fokus dari pulau jawa pembangunan ini membutuhkan kesabaran tetapi berpotensi memberikan dampak positif jangka panjang kenapa duit segitu gede tidak di bagi ke beberapa kota untuk pemerataan ekonomi ikn sudahlah biaya besar dan terfokus ke satu titik prioritaskan pekerjaan untuk rakyat pak tanpa rakyat yang bekerja negaranya bisa lumpuh</p>

kita tidak butuh ikn pak yang kita butuh sekolah-sekolah dan fasilitas kesehatan yang merata di daerah-daerah
mending dana triliunan untuk ikn digunakan untuk subsidi pupuk subsidi listrik dan buat bangun sekolah ikn itu proyek tidak berguna

- d. Tokenizer: Tahap yang dilakukan setelah proses case folding untuk memisahkan kata-kata menjadi beberapa token dalam sebuah kalimat atau tweet. Tabel 6 berikut ini merupakan hasil dari proses normalisasi.

Tabel 6. Ilustrasi Tokenizer

Tokenizer
<p>“[‘ikn’, ‘menjadi’, ‘langkah’, ‘strategis’, ‘untuk’, ‘menciptakan’, ‘pemerataan’, ‘ekonomi’, ‘dan’, ‘penduduk’, ‘menjauhkan’, ‘fokus’, ‘dari’, ‘pulau’, ‘jawa’, ‘pembangunan’ ‘ini’, ‘membutuhkan’, ‘kesabaran’, ‘tetapi’, ‘berpotensi’, ‘memberikan’, ‘dampak’, ‘positif’, ‘jangka’, ‘panjang’]”</p> <p>“[‘kenapa’, ‘duit’, ‘segitu’, ‘gede’, ‘tidak’, ‘di’, ‘bagi’, ‘ke’, ‘beberapa’, ‘kota’, ‘untuk’, ‘pemerataan’, ‘ekonomi’, ‘ikn’, ‘sudahlah’, ‘biaya’, ‘besar’, ‘dan’, ‘terfokus’, ‘ke’, ‘satu’, ‘titik’]”</p> <p>“[‘prioritaskan’, ‘pekerjaan’, ‘untuk’, ‘rakyat’, ‘pak’, ‘tanpa’, ‘rakyat’, ‘yang’, ‘bekerja’, ‘negaranya’, ‘bisa’, ‘lumpuh’]”</p> <p>“[‘kita’, ‘gak’, ‘butuh’, ‘ikn’, ‘pak’, ‘yg’, ‘kita’, ‘butuh’, ‘sekolah2’, ‘dan’, ‘fasilitas’, ‘kesehatan’, ‘yg’, ‘merata’, ‘di’, ‘daerah2’]”</p> <p>“[‘mending’, ‘dana’, ‘triliunan’, ‘utk’, ‘ikn’, ‘digunakan’, ‘utk’, ‘subsidi’, ‘pupuk’, ‘subsidi’, ‘listrik’, ‘dan’, ‘buat’, ‘bangun’, ‘sekolah’, ‘ikn’, ‘itu’, ‘proyek’, ‘gak’, ‘berguna’]”</p>

- e. Stopword: Langkah ini dilakukan untuk membersihkan kata-kata hasil normalisasi dari kata-kata yang kurang bermakna atau kata-kata umum yang tidak memberikan kontribusi signifikan dalam analisis. Tabel 7 berikut ini merupakan hasil dari proses filtering.

Tabel 7. Ilustrasi Stopwords

Stopword
<p>ikn menjadi langkah strategis menciptakan pemerataan ekonomi penduduk menjauhkan fokus pulau jawa pembangunan ini membutuhkan kesabaran tetapi berpotensi memberikan dampak positif jangka panjang kenapa duit segitu gede tidak bagi beberapa kota untuk pemerataan ekonomi ikn sudahlah biaya besar dan terfokus satu titik prioritaskan pekerjaan untuk rakyat pak tanpa rakyat bekerja negaranya lumpuh</p> <p>kita tidak butuh ikn pak kita butuh sekolah-sekolah fasilitas kesehatan merata daerah-daerah mending dana triliunan ikn digunakan subsidi pupuk subsidi listrik bangun sekolah ikn proyek tidak berguna</p>

- f. Stemming: Tahap dimana semua kata diubah menjadi bentuk dasarnya, sehingga kata-kata dengan akar yang sama dianggap serupa. Tabel 8 berikut ini merupakan hasil dari proses stemming.

Tabel 8. Ilustrasi Stemming

Stemming
<p>langkah strategis cipta rata ekonomi potensi beri dampak positif jangka panjang</p> <p>kenapa uang segitu besar tidak bagi berapa kota untuk rata ekonomi ikn sudah biaya besar fokus satu titik</p>

prioritas kerja untuk rakyat bapak tanpa rakyat kerja negara bisa lumpuh
 kita tidak butuh ikn bapak kita butuh sekolah dan fasilitas sehat rata di daerah
 mending dana triliun untuk ikn guna untuk subsidi pupuk subsidi listrik buat
 bangun sekolah ikn itu projek tidak guna

3.4 Pembobotann TF-IDF

(218, 133)	0.8838170249799399
(218, 296)	0.277685591578652
(218, 66)	0.3765078732048274
(219, 63)	0.5292164647798075
(219, 484)	0.5292164647798075
(219, 349)	0.40455589258661884
(219, 784)	0.48686328745502166
(219, 296)	0.1978851583971329
(220, 764)	0.3256144986778897
(220, 821)	0.6512289973557794
(220, 138)	0.28608431602711004
(220, 874)	0.28608431602711004
(220, 5)	0.30249082418052564
(220, 766)	0.2629606415297459
(220, 616)	0.21463924584338737
(220, 444)	0.3153862963683851

Gambar 2. Hasil Pembobotan TF-IDF

Hasil dari pembobotan kata menggunakan TF-IDF berkisaran antara 0 – 1, jika nilai menunjukan hasil yang mendekati 1 berarti kata tersebut memiliki frekuensi yang tinggi pada data. Gambar diatas menunjukan pembobotan kata pada komentar 218 hingga 220.

3.5 Analisis Hasil

- a. Hasil Prediksi Precision, Recall, Dan F1-Score

Tabel 9. Hasil Prediksi Data

	precision	recall	F1-score
Negatif	0.67	0.80	0.73
Netral	0.25	0.12	0.17
Positif	0.75	0.75	0.75
Weighted avg	0.62	0.66	0.63

Analisis Sentimen menggunakan klasifikasi Naïve Bayes mendapatkan nilai masing-masing Precision 62%, Recall 66%, dan f1-Score 63%.

- b. Prediksi Confusion Matrix

Confusion matrix merupakan sebuah matriks evaluasi yang berbentuk matrix, digunakan untuk menghitung jumlah akurasi klasifikasi terhadap kelas dengan menggunakan algoritma Naïve Bayes [18]. Berikut tabel dari hasil pengujian data testing:

Tabel 10. Tabel Confudion Matrix

Pelabelan	Prediksi		
	Negatif	Netral	Positif
Negatif	16	2	2
Netral	5	1	2
Positif	3	1	12

Hasilnya:



Gambar 4. Visualisasi Sentimen Netral

Untuk komentar yang dikategorikan sebagai netral, kata-kata yang sering muncul melibatkan istilah seperti IKN, bangun, negara, rakyat, proyek, dan lainnya. Dari rangkaian kata-kata ini, dapat diartikan bahwa komentar berlabel netral cenderung menyajikan saran atau pandangan dari perspektif alternatif terkait pembangunan IKN. Hal ini melibatkan pertimbangan terhadap jumlah kata positif dan negatif dalam konteks pembangunan IKN. Beberapa masyarakat mungkin setuju terhadap konsep pembangunan IKN, namun tidak setuju terhadap metode pelaksanaannya. Kritik mungkin mencakup persepsi bahwa proyek ini terlalu tergesa-gesa, menggunakan dana Anggaran Pendapatan dan Belanja Negara (APBN), serta menyebabkan deforestasi dengan menghilangkan jutaan pohon di Kalimantan.

3) Pelabelan Sentimen Positif



Gambar 5. Visualisasi Sentimen Positif

Untuk sentimen positif, kata-kata yang sering muncul melibatkan istilah seperti IKN, kota, harapan, moga, Jokowi, Indonesia, dan sebagainya. Hasil tersebut mencerminkan bahwa komentar positif cenderung berfokus pada aspirasi, dukungan, dan kagum terhadap proyek IKN. Banyak orang berharap bahwa IKN akan membawa perbaikan signifikan dalam perencanaan kota dan dapat menjadi teladan bagi kota-kota lain di masa depan. Meskipun mungkin IKN merupakan proyek yang besar dan memerlukan investasi negara yang substansial, namun jika dapat dijalankan dengan lancar dan mencapai targetnya, kemungkinan besar IKN akan menjadi pusat ekonomi dan pemerintahan yang mendukung kemajuan Indonesia ke depan.

3.6 Klasifikasi Naive Bayes Classifier

Dalam studi ini, peneliti juga mengembangkan sebuah program menggunakan algoritma Naive Bayes dengan menggunakan bahasa pemrograman Python. Program ini dapat secara langsung memprediksi hasil ketika komentar baru dimasukkan ke dalam sistem. Untuk melihat hasil pengujian program, dapat dilihat pada Gambar 6.

```
new_text = input("\nMasukan teks baru: ")
new_text_vec = vectorizer.transform([new_text])
predict_sentimen = naive_bayes.predict(new_text_vec)

if predict_sentimen[0]==2:
    sentimen_label = "Positif"
elif predict_sentimen[0]==1:
    sentimen_label = "Netral"
elif predict_sentimen[0]==0:
    sentimen_label = "Negatif"

print("Hasil Analisis Sentiment teks baru =",sentimen_label)
```

Masukan teks baru: Menghibur diri dibalik kecemasan tidak ada investor masuk. Memindahkan ibukota tak semudah pindah kontrakan pakde.
Hasil Analisis Sentiment teks baru = Negatif

Gambar 6. Hasil Program Naive Bayes

Dalam program, sebuah komentar baru dimasukkan dengan kalimat "Menghibur diri di balik kecemasan tidak ada investor yang masuk. Memindahkan ibukota tidak semudah pindah kontrakan pakde." Sistem mendeteksi bahwa komentar tersebut termasuk dalam kategori komentar negatif, hal itu dikarenakan banyaknya kata-kata yang cenderung menuju sentimen negative didalam komentar tersebut.

4. PENUTUP

Analisis sentimen terhadap mega proyek pembangunan Ibu Kota Negara (IKN) Nusantara, yang diinisiasi oleh Presiden Joko Widodo, memunculkan berbagai tanggapan positif dan negatif di kalangan masyarakat. Penelitian ini menggunakan algoritma Naive Bayes untuk menganalisis pandangan masyarakat melalui platform media sosial Twitter. Naive Bayes, sebagai salah satu metode pembelajaran mesin yang populer digunakan, khususnya untuk analisis sentimen publik. Dengan menggunakan bahasa pemrograman Python, peneliti menganalisis 220 komentar pada akun Twitter Presiden Joko Widodo, terdiri dari 116 komentar negatif, 70 komentar positif, dan 34 komentar netral. Hasil analisis menunjukkan bahwa program yang dikembangkan menggunakan algoritma Naive Bayes memiliki akurasi sebesar 66%. Data dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian, di mana 44 komentar diuji dapat memprediksi 20 komentar negatif, 8 komentar netral, dan 16 komentar positif. Visualisasi Wordcloud juga digunakan untuk menggambarkan inti dari setiap kategori komentar, dengan kata-kata terbesar mencerminkan signifikansi sentimen. Dalam pengujian program dengan memasukkan komentar baru di luar dataset, program mampu memprediksinya meskipun dengan tingkat akurasi yang masih perlu ditingkatkan. Dari hasil analisis sentimen, dapat disimpulkan bahwa pendapat masyarakat terkait pembangunan IKN masih kontroversial. Meskipun banyak yang tidak setuju, proyek ini diarahkan untuk jangka panjang dengan harapan dapat meningkatkan stabilitas ekonomi, kesejahteraan masyarakat, dan menjadikan Indonesia sebagai negara maju.

DAFTAR PUSTAKA

- [1] M. K. Saraswati *et al.*, "Pemindahan Ibu Kota Negara Ke Provinsi Kalimantan Timur Berdasarkan Analisis Swot," *Jurnal Ilmu Sosial dan Pendidikan (JISIP)*, vol. 6, no. 2, pp. 2598–9944, 2022, doi: 10.36312/jisip.v6i1.3086/http.
- [2] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional," *Jurnal Teknokompak*, vol. 15, no. 1, pp. 131–145, 2021.
- [3] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 10, no. 1, Jan. 2022, doi: 10.23960/jitet.v10i1.2262.
- [4] N. Dwitiyanti and N. Selvia, "Analisis Sentimen Twitter Kebiasaan New Normal," *Seminar Nasional Riset dan Teknologi dan Inovasi Teknologi (SEMNAS RISTEK)*, no. 2020, 2021.
- [5] I. Novitasari, T. B. Kurniawan, D. A. Dewi, and Misinem, "Analisis sentimen masyarakat terhadap tweet ruang guru menggunakan algoritma naive bayes classifier (NBC) [Analysis of public sentiment towards ruang guru's tweets using the Naive Bayes Classifier (NBC) algorithm]," *Jurnal Mantik*, vol. 6, no. 3, 2022.

- [6] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, "Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *JURIKOM (Jurnal Riset Komputer)*, vol. 8, no. 6, p. 219, Dec. 2021, doi: 10.30865/jurikom.v8i6.3655.
- [7] S. N. J. Fitriyyah, N. Safriadi, and E. E. Pratama, "Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 5, no. 3, 2019, doi: 10.26418/jp.v5i3.34368.
- [8] F. Nurhuda, S. W. Sihwi, and A. Doewes, "Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier," vol. 2, no. 2, 2013.
- [9] M. Pramadani, R. Putra, K. Rizky, and N. Wardani, "JUTIM (Jurnal Teknik Informatika Musirawas) PENERAPAN TEXT MINING DALAM MENGANALISIS KEPRIBADIAN PENGGUNA MEDIA SOSIAL," 2020.
- [10] R. L. Atimi and Enda Esyudha Pratama, "Implementasi Model Klasifikasi Sentimen Pada Review Produk Lazada Indonesia," *Jurnal Sains dan Informatika*, vol. 8, no. 1, 2022, doi: 10.34128/jsi.v8i1.419.
- [11] M. Azahri, N. Sulistiyowati, and M. Jajuli, "ANALISIS SENTIMEN PENGGUNA KERETA API INDONESIA MELALUI SOSIAL MEDIA TWITTER DENGAN ALGORITMA NAÏVE BAYES CLASSIFIER," 2023.
- [12] M. C. Kirana, N. P. Perkasa, M. Z. Lubis, and M. Fani, "Visualisasi Kualitas Penyebaran Informasi Gempa Bumi di Indonesia Menggunakan Twitter," *JOURNAL OF APPLIED INFORMATICS AND COMPUTING*, 2019, doi: 10.30871/jaic.v0i0.1246.
- [13] E. Febriyani and H. Februariyanti, "Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Algoritma Naive Bayes Classifier Di Twitter," *Jurnal TEKNO KOMPAK*, vol. 17, no. 1, pp. 25–38, 2023.
- [14] T. Yusnitasari, D. Ikasari, E. E. S. Pratiwi, and N. S. Ramdani, "Analisis Sentimen Terhadap Review Restoran Fish Streat Pada Aplikasi Zomato Menggunakan Stemming Nazief Adriani Dan Naive Bayes Classifier," *Prosiding Sentrinov*, vol. Vol 3, 2017.
- [15] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," 2021.
- [16] A. Tumanggor and S. Hasugian, "Penerapan Data Mining Untuk Memprediksi Tingkat Kemampuan Anak Dalam Mengikuti Mata Pelajaran Dengan Metode C4.5 Pada SDN 105351 Bakaran Batu," *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 4, no. 1, 2021.
- [17] M. I. Petiwi, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Gofood Berdasarkan Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 1, 2022, doi: 10.30865/mib.v6i1.3530.
- [18] L. Qadrini, A. Sepperwali, and A. Aina, "Decision Tree Dan Adaboost Pada Klasifikasi Penerima Program Bantuan Sosial," *Jurnal Inovasi Penelitian*, vol. 2, no. 7, 2021.